УДК 519.24
DOI:10.17853/2587-6910-2020-03-19-22

ПРИМЕНЕНИЕ БИНОМИАЛЬНЫХ МОДЕЛЕЙ С ЛАТЕНТНЫМИ ПАРАМЕТРАМИ ДЛЯ ИССЛЕДОВАНИЯ ОЦЕНОК

THE USE OF BINOMIAL MODELS WITH LATENT PARAMETERS FOR STUDY MARKS

Владимир Владимирович Братищенко
кандидат физико-математических наук, доцент
e-mail: vvb@bgu.ru
ФГБОУ ВО «Байкальский государственный университет», Иркутск, Россия

Vladimir Vladimirovich Bratischenko
Baikal State University, Irkutsk, Russia

Аннотация. *Для моделирования учебных оценок студентов предлагается использовать бета-биномиальное распределение с латентными параметрами («подготовленность студента» и «трудность задания»). Методом моментов получены уравнения, связывающие латентные переменные и статистические характеристики оценок. Разработаны численные методы решения полученных уравнений. Статистическая обработка набора оценок обосновывает применение модели для изучения методик оценивания.*

Ключевые слова: *современная теория тестирования, модели с латентными параметрами, подготовленность студента, трудность задания, бета-биномиальное распределение.*

Abstract. *The article proposes to use the beta-binomial distribution with latent parameters for the simulation of students' educational marks: student's "ability" and task "difficulty". Using the method of moments, equations are obtained that link latent variables and statistical characteristics of marks. Numerical methods for solving the obtained equations are developed. Statistical processing of a marks set justifies the use of the model to study estimation methods.*

Keywords: *Item Response Theory, models with latent variables, person ability, item difficulty, Beta-Binomial Distribution.*

Исследование статистических характеристик оценок [1] позволяет утверждать, что оценки, выставленные преподавателями по разным дисциплинам, не принадлежат одной генеральной совокупности. Но при формировании сводного итогового документа требуется, чтобы они были сопоставимы для правильного отражения итогов обучения. В современной теории тестирования (Item Response Theory, IRT) для учета и уровня подготовленности тестируемых, и уровня трудности заданий применяют латентные переменные [2]. В научных исследованиях [3, 4] предлагается применить этот подход для учета оценок преподавателей.

Первоначально IRT разрабатывалась для обработки дихотомических заданий, в которых возможны два варианта ответа: «правильно» — 1 и «неправильно» — 0. Впоследствии этот подход был распространен на политомические задания с несколькими вариантами ответа различной степени правильности (точности) [5, 6]. При этом два основных латентных параметра («трудность задания» и «подготовленность тестируемого») дополняются параметрами вариантов ответа. Такая универсальная модель позволяет разделить варианты ответа по трудности. Однако для надежности оценивания большого количества параметров требуется больше наблюдений.

В биномиальной модели [3] вероятность того, что i-й студент получит оценку $k \in \{0,\ldots,N\}$ за j-е задание, имеет следующее биномиальное распределение:

$$P\left(X_{ij} = k\right) = C_N^k p_{ij}^k q_{ij}^{N-k}$$

где X_{ij} — оценка i-го студента за j-е задание;

$$p_{ij} = \frac{e^{\theta_i}}{e^{\theta_i} + e^{\delta_j}}$$

$$q_{ij} = 1 - p_{ij} = \frac{e^{\delta_j}}{e^{\theta_i} + e^{\delta_j}}$$

где θ_i — параметр подготовленности i-го студента ($i = 1, \ldots, n$);

δ_j — параметр трудности j-го задания ($j = 1, \ldots, m$).

Все оценки полагаются независимыми в совокупности. Биномиальное распределение соответствует дискретной природе оценок и предположению, что оценка преподавателя складывается под действием многостороннего исследования выполнения задания студентом и должна быть близка к сумме бернуллиевских случайных величин. По диапазону значений подобное распределение подходит и для традиционной шкалы оценок, и для активно внедряемой в последнее время стобалльной шкалы.

Исследования оценок показали, что модель в целом проходит статистические проверки по критерию Фишера и традиционные для IRT проверки с применением Infit и Outfit статистик [2]. Однако оценки некоторых заданий не проходят проверки, что объясняется неподходящими методиками: или все задания оцениваются одинаково, или появляются выбросы, не соответствующие средним показателям успеваемости студента.

В то же время обнаружилось, что оценки имеют большую дисперсию, чем это предсказывает модель. Для устранения этого недостатка нужно использовать распределение, аналогичное биномиальному, но имеющее дополнительные параметры, влияющие на дисперсию. Таким распределением является бета-биномиальное.

Бета-биномиальная распределение экзаменационной оценки задается следующим образом:

$$P\left\{X_{ij} = x_{ij}\right\} = C_{k_j}^{x_{ij}} p_{ij}^{x_{ij}} (1 - p_{ij})^{k_j - x_{ij}}$$

где k_j — количество градаций оценки j-го задания;

p_{ij} — случайная величина, имеющая бета-распределение

$$f(x) = \frac{x^{\theta_i - 1}(1-x)^{\delta_j - 1}}{B(\theta_i, \delta_j)}$$

которое зависит от параметров подготовленности i-го студента ($\theta_i > 0$) > и трудности j-го задания ($\delta_j > 0$).

В результате усреднения по распределению p_{ij} получаем вероятности оценок γ

$$P\left\{X_{ij} = x_{ij}\right\} = \frac{\Gamma(k_j + 1)}{\Gamma(x_{ij}+1)\Gamma(k_j - x_{ij} + 1)} \frac{\Gamma(x_{ij} + \theta_i)\Gamma(k_j - x_{ij} + \delta_j)}{\Gamma(k_j - \theta_i + \delta_j)} \frac{\Gamma(\theta_i + \delta_j)}{\Gamma(\theta_i)\Gamma(\delta_j)}$$

и числовые характеристики распределения

$$M\left[X_{ij}\right] = \frac{k_j \theta_i}{\theta_i + \delta_j}$$

$$D\left[X_{ij}\right] = \frac{k_j \theta_i \delta_j (k_j + \theta_i + \delta_j)}{(\theta_i + \delta_j)^2 (1 + \theta_i + \delta_j)}$$

Особенностью параметров θ_i и δ_j в данной модели является то, что при умножении их на некоторый положительный коэффициент γ математическое ожидание не изменится, а дисперсия будет варьироваться в широких пределах – от $k_j p_{ij}\left(1 - p_{ij}\right)$ при $\gamma \to \infty$, что соответствует дисперсии биномиального распределения, до $k_j^{\,2} p_{ij}\left(1 - p_{ij}\right)$ при $\gamma \to 0$

При обработке оценок (особенно оценок текущей успеваемости) достаточно часто встречаются неполные наборы оценок (например, если некоторые студенты не выполняют какие-либо учебные задания). Для описания такого набора оценок введем следующие индикаторы: i-го студента за j-е задание,

$$I_{ij} = \begin{cases} 1, \text{если есть оценка } i-\text{го студента за } j-\text{е задание}, \\ 0, \text{если нет оценки } i-\text{го студента за } j-\text{е задание}. \end{cases}$$

Для оценки параметров модели можно воспользоваться методом моментов. Используя формулу для математического ожидания случайной величины X_{ij} и выполняя суммирования по i и j, получаем уравнения

$$\sum_{j=1}^{m} I_{ij} \frac{k_j \theta_i}{\theta_i + \delta_j} - \sum_{j=1}^{m} I_{ij} x_{ij} = 0$$

$$\sum_{j=1}^{n} I_{ij} \frac{k_j \theta_i}{\theta_i + \delta_j} - \sum_{j=1}^{n} I_{ij} x_{ij} = 0$$

которые можно применять для подбора параметров. Традиционно в IRT для этого используется метод Ньютона (касательных). В этом методе итерационная формула

$$x^{(k+1)} = x^{(k)} - \frac{f\left(x^{(k)}\right)}{f'\left(x^{(k)}\right)}$$

использует производные по параметрам

$$\frac{\partial}{\partial \theta_i}\left(\sum_{j=1}^{m} I_{ij} \frac{k_j \theta_i}{\theta_i + \delta_j} - \sum_{j=1}^{m} I_{ij} x_{ij} \right) = \sum_{j=1}^{m} I_{ij} \frac{k_j \delta_j}{(\theta_i + \delta_j)^2}$$

$$\frac{\partial}{\partial \delta_j}\left(\sum_{i=1}^{n} I_{ij} \frac{k_j \theta_i}{\theta_i + \delta_j} - \sum_{i=1}^{n} I_{ij} x_{ij} \right) = -\sum_{i=1}^{n} I_{ij} \frac{k_j \theta_j}{(\theta_i + \delta_j)^2}$$

В итоге получаем формулы

$$\theta_i^{(k+1)} = \theta_i^{(k)} - \frac{\sum_{j=1}^{m} I_{ij} \frac{k_j \theta_i^{(k)}}{\theta_i^{(k)} + \delta_j^{(k)}} - \sum_{j=1}^{m} I_{ij} x_{ij}}{\sum_{j=1}^{m} I_{ij} \frac{k_j \delta_j^{(k)}}{(\theta_i^{(k)} + \delta_j^{(k)})^2}}$$

$$\delta_j^{(k+1)} = \delta_j^{(k)} + \frac{\sum_{i=1}^{n} I_{ij} \frac{k_j \theta_i^{(k)}}{\theta_i^{(k)} + \delta_j^{(k)}} - \sum_{i=1}^{n} I_{ij} x_{ij}}{\sum_{i=1}^{n} I_{ij} \frac{k_j \theta_i^{(k)}}{(\theta_i^{(k)} + \delta_j^{(k)})^2}}$$

для определения оценок параметров численными методами. Для вычисления начальных значений положим, что в первом уравнении все параметры экзаменов одинаковы $\delta_j^{(0)} = \bar{\delta}$ в этом случае

$$\theta_i^{(0)} = \frac{\bar{\delta} \sum_{i=1}^{m} I_{ij} x_{ij}}{\sum_{i=1}^{m} I_{ij} (k_j - x_{ij})}$$

Аналогично, полагая $\theta_i^{(0)} = \bar{\theta}$

$$\delta_j^{(0)} = \frac{\bar{\theta} \sum_{i=1}^{n} I_{ij} (k_j - x_{ij})}{\sum_{i=1}^{n} I_{ij} x_{ij}}$$

Значения $\bar{\theta} = 7{,}5$ и $\bar{\delta} = 2{,}5$ обеспечивают начальные значения оценок, соответствующие самой массовой оценке — «хорошо». Вычислительный процесс обладает хорошей сходимостью при соблюдении условия вариативности оценок: отсутствуют или исключены ситуации, когда все оценки некоторого задания минимальны или максимальны. Также поиск параметров можно представить в виде процедуры минимизации неотрицательной квадратичной формы, что обеспечивает сходимость итерационных вычислений.

Обработка данных показала, что математические ожидания оценок биномиальной и бета-биномиальной моделей совпадают. Оценка дисперсии остатков соответствует вычисленному значению бета-биномиальной модели. Однако для разных заданий отклонение оценки дисперсии от теоретического значения оказывается значительным (уровни подготовленности к выполнению различных заданий отличаются и должны описываться разными латентными переменными). Применение предложенного подхода будет более обоснованным, если можно выделить задания, требующие сходных компетенций. Это позволит более точно измерять сформированность компетенций по итогам обучения.

Список литературы

1. *Братищенко, В. В.* Статистический анализ экзаменационных оценок / В. В. Братищенко. Текст: электронный // Известия Иркутской государственной экономической академии (Байкальский государственный университет экономики и права). 2011. № 3. URL: http://eizvestia.isea.ru/reader/article.aspx?id=8014.

2. *Нейман, Ю. М.* Введение в теорию моделирования и параметризации педагогических тестов / Ю. М. Нейман, В. А. Хлебников. Москва: Прометей, 2000. 168 с. Текст: непосредственный.

3. *Братищенко, В. В.* Параметрическая модель экзаменационных оценок / В. В. Братищенко. Текст: непосредственный // Качество. Инновации. Образование. 2012. № 3 (82). С. 32–35.

4. *Родионов, А. В.* Применение IRT-моделей для анализа результатов обучения в рамках компетентностного подхода / А. В. Родионов, В. В. Братищенко. Текст: электронный // Современные проблемы науки и образования. 2014. № 4. URL: www.science-education.ru/118–13858.

5. *Masters, N. G.* A Rasch model for partial credit scoring / N. G. Masters. Text: print // Psychometrika. 1982. Vol. 47. P. 149–174.

6. *Andrich, D.* A rating formulation for ordered response categories / D. Andrich. Text: print // Psychometrika. 1978. Vol. 45. P. 561–573.